

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-204376

(43) 公開日 平成9年(1997)8月5日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 13/00	3 5 1		G 0 6 F 13/00	3 5 1 C
H 0 4 L 12/46			H 0 4 L 11/00	3 1 0 C
12/28		9466-5K	11/20	B
12/66			13/00	3 0 5 Z
29/06				

審査請求 未請求 請求項の数10 O L (全 27 頁)

(21) 出願番号 特願平8-11836

(22) 出願日 平成8年(1996)1月26日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 飯塚 史之

神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(72) 発明者 鳥居 悟

神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(74) 代理人 弁理士 井桁 貞一

(54) 【発明の名称】 通信宛先管理方法および装置

(57) 【要約】

【課題】 本発明は多階層のネットワークプロトコルを用いたプロセス間通信技術に関し、ゲートウェイの負荷を低減し相手先のプロセッサに負荷を分散させること、および内部ネットワークに接続される並列／分散システムを1つのシステムとして扱うことを可能にする通信宛先管理方法および装置を提供することを目的とする。

【解決手段】 本発明通信宛先管理方法および装置は以下のように構成する。

(1) 宛先ポートを読み出し、対応する物理アドレスを書き換える。

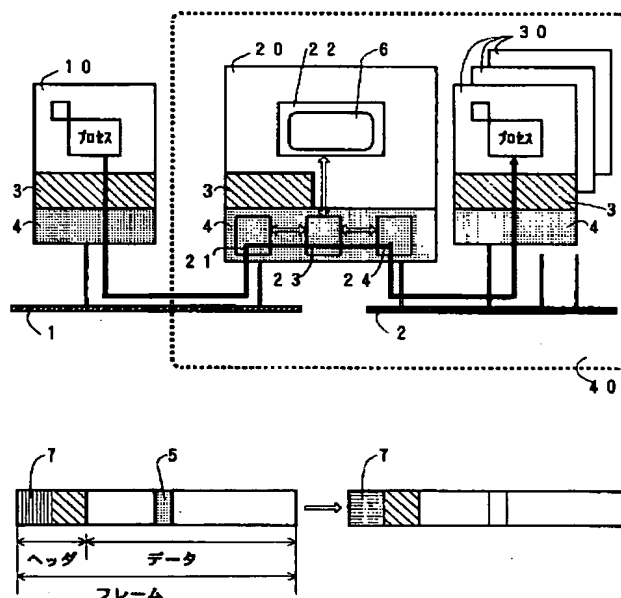
(2) ポートとプロセッサの物理アドレスの対応表記憶手段を設け、クライアントサーバ方式でポートを管理する。

(3) フラグメント記憶手段を設け、フラグメントの順序性を保証する。

(4) 上記(1)と(3)を下位プロトコルで行うデータリンクブリッジ手段を設ける。

(5) 複数のネットワークデータリンク処理手段を設ける。

第1の発明の原理図



【特許請求の範囲】

【請求項1】 外部のネットワークを介してホストコンピュータもしくはプロセッサエレメントもしくは他のゲートウェイに接続され、且つ、

内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイを経由して、

上位プロトコルと下位プロトコルから成る階層構造のネットワークプロトコルによってプロセス間通信を行う場合、

通信の相手先プロセスに固有の前記上位プロトコルの宛先ポート情報を、前記下位プロトコルのフレームのデータ部分から読み出して、該宛先ポートから宛先の前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスを調べ、前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えることを特徴とする並列あるいは分散システムにおける通信宛先管理方法。

【請求項2】 内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイにおいて、

前記ホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を設け、

前記プロセスの実行開始時に、該プロセスの要求に応じて前記ポート情報を該プロセスに割り付け、該ポート情報を前記対応表に登録し、該プロセスの実行終了時に、該プロセスの要求に応じて前記対応表から該ポート情報を削除すると共に、

前記プロセスの要求に応じて、前記対応表を検索し前記ポート情報を該プロセスに通知することを特徴とする並列あるいは分散システムにおける通信宛先ポート管理方法。

【請求項3】 前記プロセス間通信中にパケットのフラグメント化が発生し、該通信の宛先ポート情報を含む先頭フラグメントが他のフラグメントより遅れて到着する場合に、

前記フラグメント化発生時に全フラグメントに付与される該パケットの識別子と共にフラグメントの順番を示す識別子を参照することによって、前記先頭フラグメントを特定し、該先頭フラグメントに含まれる前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えてから全フラグメントを送信することを特徴とする請求項1に記載の並列あるいは分散システムにおけるフラグメントの通信宛先管理方法。

【請求項4】 前記ゲートウェイにおいて、前記下位プロトコルの内のデータリンク層に、

外部ネットワークとのインタフェースを持つ外部ネットワークデータリンク処理部と、内部ネットワークとのイ

ンタフェースを持つ内部ネットワークデータリンク処理部と、前記外部ネットワークデータリンク処理部と前記内部ネットワークデータリンク処理部の橋渡しをするデータリンクブリッジとを設け、

データリンクブリッジにより前記外部ネットワークからのパケットの宛先アドレスを書き換えて前記内部ネットワークに転送することを特徴とする請求項1～3のいずれかに記載の並列あるいは分散システムにおける通信宛先管理方法。

10 【請求項5】 前記ゲートウェイ内の前記データリンク層に、複数の前記内部データリンク処理部を設け、複数の内部ネットワーク用のデータリンクを接続し、複数クラスタ内のプロセス間通信および複数クラスタと外部ネットワークとのプロセス間通信を行うことを特徴とする請求項4に記載の並列あるいは分散システムにおける通信宛先管理方法。

【請求項6】 上位プロトコルと下位プロトコルから成る階層構造のネットワークプロトコルによってプロセス間通信を行うためのゲートウェイであって、

20 外部のネットワークを介してホストコンピュータもしくはプロセッサエレメントもしくは他のゲートウェイを接続するためのインタフェースを持つ外部ネットワークデータリンク処理手段と、

内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントを接続するためのインタフェースを持つ内部ネットワークデータリンク処理手段と、

30 前記内部のネットワークに接続されるホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を格納しておく対応表記憶手段と、

前記対応表記憶手段、前記外部ネットワークデータリンク処理手段、および前記内部ネットワークデータリンク手段に接続され、前記プロセス間通信の宛先を管理する宛先管理手段とを備え、

40 前記外部ネットワークデータリンク処理手段によって得られる通信の相手先プロセスに固有の前記上位プロトコルの宛先ポート情報を、前記下位プロトコルのフレームのデータ部分から読み出し、

前記記憶手段に格納された前記対応表から前記宛先ポートに対応する宛先のホストコンピュータもしくはプロセッサエレメントのアドレスを読み出し、

前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換え、

前記内部ネットワークデータリンク手段を経由して前記内部のホストコンピュータもしくはプロセッサエレメントに送信することを特徴とする並列あるいは分散システムにおける通信宛先管理装置。

3

【請求項 7】 内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメントに接続されるゲートウェイであって、

前記ホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメントのアドレスとの対応表を格納しておく対応表記憶手段と、

前記ホストコンピュータもしくはプロセッサエレメントに備えられたポート管理クライアント手段を介して前記プロセスから発生する要求を処理するポート管理サーバ手段とを備え、

前記ポート管理サーバ手段によって、前記ポート情報を前記プロセスに割り付け、該ポート情報を前記対応表記憶手段に登録し、前記対応表記憶手段から該ポート情報を削除すると共に、

前記プロセスの要求に応じて前記対応表記憶手段を検索し前記ポート情報を該プロセスに通知することを特徴とする並列あるいは分散システムにおける通信宛先管理装置。

【請求項 8】 前記対応表記憶手段、前記外部ネットワークデータリンク処理手段、および前記内部ネットワークデータリンク手段に接続され、フラグメントを格納するフラグメント記憶手段を備えたフラグメント処理手段を備え、

前記プロセス間通信中にパケットのフラグメント化が発生し、該通信の宛先ポート情報を含む先頭フラグメントが、他のフラグメントより遅れて到着する場合に、

前記フラグメント化発生時に全フラグメントに付与される該パケットの識別子と共にフラグメントの順番を示す識別子を参照することによって、前記先頭フラグメントを特定し、該先頭フラグメントに含まれる前記下位プロトコルのフレームのヘッダ部分の宛先アドレスを書き換えてから全フラグメントを送信することを特徴とする請求項 6 に記載の並列あるいは分散システムにおけるフラグメントの通信宛先管理装置。

【請求項 9】 前記下位プロトコルの内のデータリンク層に、

前記外部ネットワークデータリンク処理手段と、前記内部ネットワークデータリンク処理手段と、

前記宛先管理手段および前記フラグメント処理手段とから構成されるデータリンクブリッジ手段とを備え、

前記データリンク層で、前記外部ネットワークからのパケットを該宛先アドレスに応じて前記内部ネットワークに転送することを特徴とする請求項 6～8 のいずれかに記載の並列あるいは分散システムにおける通信宛先管理装置。

【請求項 10】 前記データリンク層に複数の前記内部ネットワークデータリンク処理手段を備え、

複数の内部ネットワークを接続し、外部ネットワークと

4

のプロセス間通信を行うことを特徴とする請求項 9 に記載の並列あるいは分散システムにおける通信宛先管理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は並列システムあるいは分散システムにおける通信宛先管理方法および装置に関わり、特に多階層のネットワークプロトコルを用いたプロセス間通信技術に関する。

10 【0002】

【従来の技術】 近年、大規模な分散システムの開発が本格化してきており、大量のデータ処理を高速に行う並列システムの開発や、クライアントサーバ方式の様々なアプリケーションの増加に伴い、プロセス間通信を伴うメッセージ連携のソフトウェア開発が盛んになってきている。

20 【0003】 メッセージ通信を効率的に行う研究用システムの開発や、業務用システムにおける並列システムや様々なクライアントサーバシステムの導入、更にマルチベンダーの異機種間接続などが進んでおり、速度が速くて効率良く使えるプロセス間通信が望まれている。

【0004】 こうしたプロセス間通信では、並列システムの内部に多数のプロセッサエレメントやホストコンピュータを備え、ローカルエリアネットワーク (LAN) などの内部ネットワークを介して接続し、ワイドエリアネットワーク (WAN) や LAN などの外部ネットワークを介して他の並列システムやホストコンピュータを接続し、且つそれぞれのネットワークが階層構造をしたプロトコルを持つネットワーク間通信を行う。

30 【0005】 階層化プロトコルには、一般に二つのモデルが知られている。一つは図 17 に示す ISO (国際標準化機構) の 7 層参照モデルであり、1 層の物理ハードウェア接続から 7 層のアプリケーションまでの 7 層の概念層から成る。もう一つは図 18 に示す TCP/IP Internet 階層モデルであり、最下層のハードウェア層およびネットワークインタフェース層からアプリケーションまでの 4 層の概念層 (合計 5 層) から成る。
(TCP/IP によるネットワーク構築 82p～85p 共立出版 1990 年 7 月)

40 本発明は、これらのモデルの階層間で渡されるオブジェクトをパケットとして転送するときの宛先アドレスに関わる。宛先アドレスは TCP/IP (Transmission Control Protocol/Internet Protocol) の通信プロトコルにおいては、TCP や UDP (User Datagram Protocol) のトランスポート層などの上位プロトコルでは宛先ポート情報として、IP (Internet Protocol) 層などの中位層プロトコルでは IP (Inter Process) アドレスとして、データリンク層などの下位プロトコルではホストコ

50

ンピュータもしくはプロセッサエレメントの（物理）アドレスとして定義される。一方、外部ネットワークに接続し通信を行うための機器として、IPルータやブリッジやゲートウェイがあり、ネットワークとネットワークを接続している。またハブは1台で複数のネットワークを接続することができ、複数のクラスタ接続を可能にしている。

【0006】従来のプロセス間通信では、図19に示すようにIPルータやブリッジを用いてIPアドレスによりパケット転送をするものや、ゲートウェイ上にネットワークサーバを設けてパケット転送をするものが知られている。

【0007】

【発明が解決しようとする課題】従来のIPルータやブリッジにおいては、システムを構成する全てのプロセッサエレメントにIPアドレスを付与する必要がある、プロセスのユーザも目的のプロセスがどのIPアドレスのプロセッサで実行されているかを認識しなければならないという問題があった。

【0008】並列システムや分散システムが大規模になるに従いユーザが目的のプロセスとIPアドレスの対応を認識することは困難になり、またシステムの大規模化に伴いIPアドレスが不足することが予想される。

【0009】一方、ゲートウェイ上でネットワークサーバを動作させる場合は、外部ネットワークデータリンク層、ネットワーク層、トランスポート層の各層のプロトコル処理を順次行ってから宛先ポートの情報を得るために、ゲートウェイの負荷が増大するという問題があった。

【0010】更に内部ネットワークに向けてプロセス間通信用パケットを送出するためには、内部ネットワーク用プロトコル処理であるチェックサムを取るとか、データコピーをするなどの処理ステップのために通信速度が遅くなってしまうという問題があり、システムの規模が大きくなればそれだけゲートウェイに負荷が集中してしまうという問題があった。

【0011】本発明はこのような点にかんがみてゲートウェイの負荷を低減し通信の相手先のプロセッサに負荷を分散させること、および内部ネットワークに1つのプロセッサアドレスを割り当てて並列システムもしくは分散システムを1つのシステムとして扱うことを可能にすると共に、複数クラスタの通信宛先管理を可能とし更に大規模な並列システムもしくは分散システムを構築するという通信宛先管理方法および装置を提供することを目的とする。

【0012】

【課題を解決するための手段】上記の課題は下記の如くに構成された本発明の並列あるいは分散システムにおける通信宛先管理方法および装置によって解決される。

【0013】（1）本発明の第1は、請求項1および請

求項6にそれぞれ記載の通信宛先管理方法および装置である。図1は第1の発明の原理図である。図1において1は外部ネットワーク、2は内部ネットワーク、3は上位プロトコル、4は下位プロトコルを示す。上位プロトコル3と下位プロトコル4の間には一般にIPプロトコルなどの中位層プロトコルがあるが図示していない。図のフレームの中にあるデータ部分にある5は宛先ポート情報、同じくヘッダ部分にある7は宛先アドレスを示す。

10 【0014】10は外部ネットワーク1に接続されるホストコンピュータもしくはプロセッサエレメントである。10は図示していない他のゲートウェイを経由して外部ネットワークに接続される場合もある。20はゲートウェイであり、本発明の通信宛先管理装置である。

【0015】30は内部ネットワーク2に接続されるホストコンピュータもしくはプロセッサエレメントであって複数ある。点線で囲った40は一つの並列システムあるいは分散システムである。

20 【0016】6は宛先ポート情報5と宛先アドレス7の対応を示すポート情報に関するアドレス対応表である。ポート・アドレス対応表6は、内部ネットワーク2に接続されるホストコンピュータもしくはプロセッサエレメントで実行されるプロセスがプロセス間通信で使用するポート情報と、該プロセスを実行する前記内部のホストコンピュータもしくはプロセッサエレメントのアドレスとの対応を示す。

【0017】21は外部ネットワーク1を接続するためのインタフェースを持つ外部ネットワークデータリンク処理手段、22はポート・アドレス対応表6を格納する対応表記憶手段、24は内部ネットワーク2を接続するためのインタフェースを持つ内部ネットワークデータリンク処理手段である。

【0018】23は外部ネットワークデータリンク処理手段21、対応表記憶手段22、および内部ネットワークデータリンク手段24に接続され、プロセス間通信の宛先を管理する宛先管理手段である。

【0019】第1の発明の方法はゲートウェイである第1の発明の装置20を経由して、上位プロトコル3と下位プロトコル4から成る階層構造のネットワークプロトコルによってプロセス間通信を行う場合に、通信の相手先プロセスに固有の上位プロトコルの宛先ポート情報5を、下位プロトコル4のフレームのデータ部分から読み出して、宛先ポートから宛先の内部のホストコンピュータもしくはプロセッサエレメント30のアドレスを調べ、下位プロトコルのフレームのヘッダ部分の宛先アドレス7を書き換えることを特徴とする。

【0020】第1の発明の装置20は、前記第1の発明の方法に記載のプロセス間通信を行うためのゲートウェイであって、外部ネットワークデータリンク処理手段21と、対応表記憶手段22と、宛先管理手段23と、内

部ネットワークデータリンク処理手段 2 4 とを備える。

【0 0 2 1】宛先管理手段 2 3 では、まず相手先プロセスの上位プロトコル 3 の宛先ポート情報 5 を読み出す。宛先ポート情報 5 は外部ネットワークデータリンク処理手段 2 1 によって得られる、下位プロトコル 4 のフレームのデータ部分にある。

【0 0 2 2】宛先管理手段 2 3 では、次に対応表記憶手段 2 2 に格納された対応表 6 から宛先ポート情報 5 に対応する宛先のホストコンピュータもしくはプロセッサエレメントのアドレス 7 を読み出す。

【0 0 2 3】読み出された宛先アドレス 7 が下位プロトコル 4 のフレームのヘッダ部分の宛先アドレスに書き換えられる。続いて下位プロトコル 4 のフレームが内部ネットワークデータリンク手段 2 4 を経由して内部のホストコンピュータもしくはプロセッサエレメント 3 0 に送信される。

【0 0 2 4】第 1 の発明によって、該ゲートウェイに接続される内部ネットワークを介して接続されるホストコンピュータやプロセッサエレメントのポート情報と該アドレスを管理しておけば、該ゲートウェイに一つの IP アドレスを割り付けるだけで内部ネットワークでのプロセス間通信および外部ネットワークとのプロセス間の通信が可能になると共に、上位プロトコルによる処理を実行しなくて済み該ゲートウェイの負荷を低減することが可能になる。

【0 0 2 5】(2) 本発明の第 2 は、請求項 2 および請求項 7 にそれぞれ記載の通信宛先管理方法および装置であり、前記第 1 の発明で使用される対応表に関するものである。図 2 は第 2 の発明の原理図である。図 2 の 2 5 は内部ネットワークに接続されるホストコンピュータもしくはプロセッサエレメント 3 0 に備えられたポート管理クライアントに対応するポート管理サーバである。図 2 のその他の符号は図 1 と同じである。

【0 0 2 6】ポート情報に関するアドレス対応表 6 には、内部のホストコンピュータもしくはプロセッサエレメント 3 0 で実行されるプロセスについてのポート情報と、そのプロセスが実行されるホストコンピュータもしくはプロセッサエレメント 3 0 のアドレスが示される。

【0 0 2 7】第 2 の発明の方法は、内部のネットワークを介して複数のホストコンピュータもしくはプロセッサエレメント 3 0 に接続されるゲートウェイ 2 0 において、ホストコンピュータもしくはプロセッサエレメント 3 0 で実行されるプロセスがプロセス間通信で使用するポート情報 5 と、該プロセスを実行する前記ホストコンピュータもしくはプロセッサエレメント 3 0 のアドレスとの対応表 6 を設け、対応表 6 を検索すること、およびプロセスの要求に応じて、ポート情報をプロセスに割り付けること、対応表 6 に登録すること、および対応表 6 から削除することを特徴とする。

【0 0 2 8】第 2 の発明の装置は前記第 2 の発明の方法

に記載のゲートウェイであって、対応表記憶手段 2 2 とポート管理サーバ手段 2 5 とを備え、前記第 2 の発明の方法を実現する。

【0 0 2 9】一つの並列システムあるいは分散システム 4 0 中のプロセスから、プロセスの実行開始時にポートの割付要求が発行されると、ポート管理クライアントからポート管理サーバ 2 5 にそのポートの発行および登録を依頼する。ポート管理サーバ 2 5 は、対応表記憶手段 2 2 に格納されているポート・アドレス対応表 6 を参照し、未使用のポートを選び、その対応関係を対応表 6 に登録すると共に選ばれたポートをポート管理クライアントに通知する。ポート管理クライアントはプロセスにそのポートを割り付ける。

【0 0 3 0】ポートを開放する場合や、プロセスの終了時には、ポート・アドレス対応表 6 のエントリはポート管理サーバ 2 5 によって削除される。またポート管理サーバ 2 5 はプロセスの要求に応じて対応表を検索しポート情報をプロセスに通知することもある。

【0 0 3 1】第 2 の発明によって対応表 6 を常に更新することができ、また検索することができるので前記第 1 の発明による通信宛先管理を維持することが可能になる。

(3) 本発明の第 3 は、請求項 3 および請求項 8 にそれぞれ記載の通信宛先管理方法および装置であり、前記第 1 の発明のプロセス間通信で使用されるバケットのフラグメント化が発生した場合に関するものである。図 3 は第 3 の発明の原理図である。

【0 0 3 2】フラグメントは、上位層バケットから中位層データグラムへの変換の際に、物理フレームに載るように中位層データグラムを分解するときに発生する。フラグメント化発生時には、該バケットの識別子と共にフラグメントの順番を示す識別子が全フラグメントに付与される。

【0 0 3 3】また物理層の中には転送の順序性が保証されていない物理層があり、これを使用する場合は宛先ポートを含む先頭のフラグメントが先頭以外のフラグメントより遅れて到着することがある。

【0 0 3 4】図 3 の 2 6 がフラグメントを格納するフラグメント記憶手段であり、2 7 がフラグメント記憶手段 2 6 を備えたフラグメント処理手段である。フラグメント処理手段 2 7 は、対応表記憶手段 2 2、外部ネットワークデータリンク処理手段 2 1、および内部ネットワークデータリンク手段 2 4 に接続される。図 3 のその他の符号は図 1 と同じである。

【0 0 3 5】第 3 の発明の方法は、プロセス間通信中にバケットのフラグメント化が発生し該通信の宛先ポート情報を含む先頭フラグメントが他のフラグメントより遅れて到着する場合に、ゲートウェイである通信宛先管理装置 2 0 においてフラグメント化発生時に付与されるバケット識別子および順番識別子を参照することによ

10

20

30

40

50

て、先頭フラグメントを特定し先頭フラグメントに含まれる宛先アドレス 7 を書き換えてから全フラグメントを送信することを特徴とする。

【0036】第3の発明の装置は前記第3の発明の方法に記載のゲートウェイであって、フラグメント処理手段 27 を備え、前記第3の発明の方法を実現する。第3の発明によって、パケット通信の過程でフラグメントが発生し先頭フラグメントが他のフラグメントより後で到着しても、前記第1の発明による通信宛先管理が可能になる。

【0037】(4) 本発明の第4は、請求項4および請求項9にそれぞれ記載の通信宛先管理方法および装置であり、前記第1から第3までの発明のゲートウェイに関するものである。図4は第4の発明の原理図である。

【0038】図4の28が宛先管理手段23およびフラグメント処理手段27とから構成されるデータリンクブリッジ手段である。図4のその他の符号は図1と図3と同じである。

【0039】第4の発明の方法は、ゲートウェイ20において下位プロトコル4の内のデータリンク層に、外部ネットワークデータリンク処理部と、内部ネットワークデータリンク処理部と、データリンクブリッジとを設け、データリンクブリッジにより、外部ネットワーク1からのパケットの宛先アドレスを書き換えて、内部ネットワーク2に転送することを特徴とする。

【0040】第4の発明の装置は、前記第4の発明の方法に記載のゲートウェイ20であって、下位プロトコル4の内のデータリンク層に、外部ネットワークデータリンク処理手段21と内部ネットワークデータリンク処理手段24とデータリンクブリッジ手段28とを備え、前記第4の発明の方法を実現する。

【0041】第4の発明によって外部ネットワーク1から内部ネットワーク2に転送されるパケットは、下位プロトコル4であるデータリンク層で宛先アドレスが書き換えられることになり、上位プロトコル3の処理を受けることがなくデータリンク層でデータリンクブリッジを形成することになる。

【0042】これにより下位プロトコル4だけで外部ネットワーク1および内部ネットワーク2のプロセス間通信が可能になり、ゲートウェイ20における負荷を軽減することができる。

【0043】(5) 本発明の第5は、請求項5および請求項10にそれぞれ記載の通信宛先管理方法および装置であり、前記第4の発明のゲートウェイを拡張したものである。図5は第5の発明の原理図である。

【0044】図5では複数の内部ネットワークデータリンク処理部および複数の内部ネットワークデータリンク処理手段24が設けられ、複数の内部ネットワーク2がゲートウェイであり本発明の通信宛先管理装置20に接続できる。図5のその他の符号は図1、図3、図4と同

じである。

【0045】第5の発明の方法は、ゲートウェイにおいて、下位プロトコル4の内のデータリンク層に複数の内部データリンク処理部を設け、複数の内部ネットワークを接続し、複数クラスタ内のプロセス間通信および複数クラスタと外部ネットワークとのプロセス間通信を行うことを特徴とする。

【0046】第5の発明の装置は前記第5の発明の方法に記載のゲートウェイであって、前記データリンク層に複数の前記内部ネットワークデータリンク処理手段を備え、前記第5の発明の方法を実現する。

【0047】第5の発明によって該ゲートウェイを一つのハブとしてシステムを構成することが可能になり、該ゲートウェイに一つのポートを割り付けるだけで複数のクラスタ間および外部ネットワークとのプロセス間の通信が可能になる。

【0048】

【発明の実施の形態】本発明の実施の形態について図面を用いて詳細に説明する。図6はプロセス間通信にTCP/IPを使用した場合のデータフォーマットおよび本発明の実施の形態であるシステム構成図である。

【0049】TCP/IPでプロセス間通信を行う場合、データはTCPのポートを宛先にして相手先プロセスに送られる。ユーザデータはトランスポート層でTCPヘッダを、ネットワーク層でIPヘッダを、データリンク層で物理ヘッダをそれぞれ付与されて物理フレームを形成する。物理フレームでは物理ヘッダとIPヘッダの長さが決まっているので物理フレーム内のTCPの宛先ポートの位置を特定することができる。

【0050】図6の1は外部のネットワーク、2は内部のネットワークであり、それぞれローカルエリアネットワーク(LAN)を構成する。図6の11はワークステーションであり図1のホストコンピュータもしくはプロセッサエレメント10に対応する。

【0051】図6の60はゲートウェイを兼ねるプロセッサエレメントであり、図1のゲートウェイである本発明通信宛先管理装置20に対応する。31はプロセッサエレメントであり、図1のホストコンピュータもしくはプロセッサエレメント30に対応する。40はゲートウェイ60および複数のプロセッサエレメント31で構成される並列システムである。

【0052】図6の61と64はそれぞれ外部LANデータリンクと内部LANデータリンクであり、図1の外部ネットワークデータリンク処理手段21と内部ネットワークデータリンク処理手段24に対応する。62はポート・PE対応表であり、図1のポート・アドレス対応表6に対応する。

【0053】図6の68はデータリンクブリッジであり、宛先管理手段23とフラグメント処理手段27から構成される。69はプロセッサエレメント31にあるポ

ート管理クライアントであり、ゲートウェイ60にあるポート管理サーバ25と共に並列システム内でのプロセスで使用されるポートを管理する。

【0054】図6において、ワークステーション11で実行されるプロセスAから並列システム40内のプロセスサ要素31で実行されるプロセスBへ、TCP/IPで通信を行う場合、プロセスAは、プロセスBのトランスポート層のポートを宛先に通信を行う。

【0055】送信データは、ワークステーション11においてプロセスAからトランスポート層プロトコル処理、ネットワーク層プロトコル処理、およびデータリンク処理を通して、外部LANに送出される。

【0056】並列システム40のゲートウェイ60では、外部LANデータリンク61においてこの送信データを受取り外部LAN用のデータリンク処理を行い、データリンクブリッジ68へこの物理フレームを渡す。

【0057】データリンクブリッジ68内の宛先管理手段23ではトランスポート層やネットワーク層のそれぞれのプロトコル処理を行わずに、物理フレーム上の宛先ポートの位置を直接読み、ポートがどのプロセスサ要素に対応付けられているかをポート・PE対応表62から判断し、物理フレームにある物理ヘッダの宛先アドレスを書き換えてから、その物理フレームを内部LANデータリンク64に渡す。

【0058】内部LANデータリンク64では、内部LAN用のデータリンク処理を行い、内部LANにパケットを送出する。プロセスBを実行するプロセスサ要素31ではパケットを受け取り、データリンク処理、ネットワーク層プロトコル処理、およびトランスポート層プロトコル処理を行い、プロセスBに渡す。

【0059】ポート・アドレス対応表62は、並列システム40の中のポート管理クライアント69およびポート管理サーバ25によって管理される。図7にポート管理クライアントサーバの詳細を示す。

【0060】ポート管理クライアント69はプロセスサ要素31の中にあり、プロセスとのインタフェースを持つインタフェース部と、ポート管理サーバ25の通信部と通信を行う通信部とから成る。

【0061】インタフェース部は既存のポート関連の機能インタフェースを通して、プロセスからのポートの割り付け要求や削除要求を受け付け、ポートの割り付けや削除を行う。ポート管理クライアント69の通信部は、ポート割り付け要求や削除要求をポート管理サーバ25に送ったり、ポート管理サーバからの新規番号の通知を受け取ったりする。

【0062】ポート管理サーバ25はゲートウェイ60の中にあり、ポート管理クライアント69の通信部と通信を行う通信部と、ポート・PE対応表62の検索、登録および削除を行う対応部とから成る。

【0063】ポート管理サーバ25の通信部は、ポート

管理クライアントからのポート割り付け要求や削除要求を受け取ったり、ポート管理サーバへ新規番号を通知したりする。対応部は、ポート管理クライアントからの要求に従い、ポート・PE対応表62のエントリを検索したり、登録したり、削除したりする。

【0064】ポート管理クライアントサーバは以下の手順で実行される。まず並列システム40内のプロセス（図7のプロセスB）がポートの割り付け要求を発行すると、ポート管理クライアントが、ゲートウェイ60のポート管理サーバ25にポートの発行および登録を依頼する。

【0065】依頼を受けたポート管理サーバ25は、ポート・PE対応表62を参照し、未使用のポートを選び、その対応関係をポート・PE対応表62に登録すると共に選ばれたポートをポート管理クライアント69に通知する。ポート管理クライアント69はプロセスBにそのポートを割り付ける。

【0066】ポートを開放する場合や、プロセスの終了時には、ポート・PE対応表62のエントリはポート管理サーバ25によって削除される。またポート管理サーバ25はプロセスBの要求に応じて対応表を検索することもある。

【0067】UNIXアプリケーションではsocket等により、トランスポート層プロトコルであるbind、connectなどの既存のインタフェースを使ってポートを割り付ける。

【0068】図8はフラグメント宛先管理構成図である。図8の(A)はTCP/IPプロトコルにおけるフラグメントの例であり、図3の第3の発明の原理図

(A)に対応する。IはTCP/IPのパケット識別子であるIP-IDENTIFICATION (IP-ID)であり、Fは同じく順番識別子であるFLAGMENT-OFFSET (オフセット)である。IとPは共に発信元のネットワーク層 (IP層) で付与される。Pはトランスポート層 (TCP層) の宛先ポートである。

【0069】物理フレームの(a)が先頭フラグメントであり、(b)が先頭以外のフラグメントである。先頭フラグメント(a)には宛先ポートが含まれる。物理フレームの転送の場合は、フラグメントの順序性が保証されていないハードウェアプロトコルがあり、(b)が先に到着する場合がある。

【0070】そのためにフラグメント化する時に全フラグメントには、パケット固有の識別子であるIP-IDと、フラグメントの順番を識別するためのオフセットが付与される。オフセットによりフラグメントの順番を知ることができる。

【0071】図8の(B)はフラグメント宛先管理構成図である。66はフラグメントバッファであり図3のフラグメント記憶手段26に対応する。27は図3乃至図6に示されるフラグメント処理手段27と同じである。

フラグメント処理手段27は、データリンク層プロトコルで外部LANデータリンク61および内部LANデータリンク64に接続される。

【0072】またフラグメント処理手段27は、ポート・PE対応表62に接続され、先頭フラグメントに含まれる宛先ポートに対応するプロセッサアドレス（PEアドレス）を読み出して、物理ヘッダの宛先アドレスを書き換える。

【0073】図8の67はIP-ID・PE対応表である。IP-ID・PE対応表67はフラグメント処理手段27が、フラグメントの物理ヘッダの宛先アドレスを書き換えるために、IP-IDとPEアドレスの対応を記憶しておくためにフラグメント処理手段27の中に備えられる。

【0074】図9はフラグメント処理手段27におけるフラグメント処理フロー図である。ステップs1では、オフセットにより先頭フラグメントかどうかを判定する。先頭フラグメントであればステップs2に進み、先頭フラグメントでなければステップs3に進む。

【0075】ステップs2では、先頭フラグメントに含まれる宛先ポートにより、対応するPEアドレスをポート・PE対応表62から読み出す。続いてIP-IDと読み出したPEアドレスの対応をIP-ID・PE対応表67に登録し、ステップs5に進む。

【0076】ステップs3では、IP-ID・PE対応表67を参照しIP-IDと宛先PEが既に登録済みかどうかを判定する。IP-IDと宛先PEが既に登録済みであれば、ステップs5に進み、未だ登録してなければステップs4に進む。

【0077】ステップs4では、そのフラグメントをフラグメントバッファ66に格納し、先頭フラグメント待ちのためにステップs1に戻る。ステップs5では、IP-ID・PE対応表67を読み出し、各フラグメントの物理フレームの物理ヘッダ部分にある宛先のプロセッサエレメントのPEアドレスを書き換え、ステップs6に進む。

【0078】ステップs6では、フラグメントバッファ66に先頭待ちの同じIP-IDを持つフラグメントがあるかどうかを判定する。同一IP-IDを持つフラグメントがあれば、ステップs7に進み、同一IP-IDを持つフラグメントがなければフラグメント処理を終了する。

【0079】ステップs7では、フラグメントバッファ66から処理待ちのフラグメントを取り出しステップs5に戻る。図10にデータリンクブリッジの構成図を示す。図10の(A)は、通信宛先管理がデータリンク層で行われることを示している。図の矢印(1)は、外部LANからのパケットが通信宛先管理により物理アドレスを書き換えられてトランスポート層やネットワーク層を経由することなしに内部LANに送出され、データリ

ンク層でブリッジしていることを示している。

【0080】矢印(2)は、外部LANからのパケットが通信宛先管理により、ゲートウェイを兼ねたプロセッサエレメント60に宛てたものと判定され、上位層であるネットワーク層やトランスポート層に渡されることを示している。

【0081】矢印(3)は、内部LAN2からのパケットが外部LANに接続されるワークステーション11や他のホストコンピュータやプロセッサエレメントが宛先であるので通信宛先管理を経由せず、外部LANに渡されることを示している。

【0082】矢印(4)も、パケットの宛先が外部LANに接続されるワークステーション11や他のホストコンピュータやプロセッサエレメントパケットであるので、プロセッサエレメント60から通信宛先管理を経由せず、外部LANに渡されることを示している。

【0083】矢印(5)と(6)はそれぞれ内部LANからプロセッサエレメント60へのパケットおよびプロセッサエレメント60から内部LANへのパケットであり、やはり通信宛先管理を経由しないで、すなわちデータリンクブリッジを通さないで、それぞれプロセッサエレメント60の上位層および内部LANに渡されることを示している。

【0084】図10の(B)は、通信宛先管理手段23およびフラグメント処理手段27から構成されるデータリンクブリッジ68が、データリンク層で外部LANデータリンク61および内部LANデータリンク64に接続され、またポート・PE対応表62を参照して通信の宛先管理をしていることを示している。

【0085】図11はデータリンクブリッジの処理フロー図である。まずステップs10で外部LANデータリンクからパケットを受け取りステップs11に進む。この時のパケットのフォーマットは物理フレームである。

【0086】ステップs11では、その物理フレームがフラグメント化しているかどうかを判定する。フラグメント化していなければ、ステップs12に進み、フラグメント化している場合はステップs13に進む。

【0087】ステップs12では、物理フレームの宛先ポート位置を算定し、その宛先ポートに基づいてポート・PE対応表62を読み出し、宛先の物理アドレスを書き換える。次にステップs14に進む。

【0088】ステップs13では、図9のフラグメントの処理フロー図に示す処理を行い、宛先の物理アドレスを書き換える。次にステップs14に進む。ステップs14では、自ノードに宛てたものかどうかを判定する。宛先が自ノードでなければステップs15へ進み、宛先が自ノードであればステップs16へ進む。

【0089】ステップs15では、内部LANにパケットを送出すべく内部LANデータリンクへ物理アドレスを書き換えた物理フレームを渡して処理を終了する。ス

ステップ16では、データリンク処理を行い、自ノードすなわちプロセッサエレメント60の上位層にデータを渡して処理を終了する。

【0090】図12は複数の内部LANデータリンク接続図である。図12では複数の内部LANデータリンク64がデータリンク層に設けられ、複数の内部LAN2がゲートウェイであるプロセッサエレメント60に接続できることを示している。

【0091】図13は複数のクラスタの通信宛先管理システム構成図である。並列システム40が複数の独立した内部ネットワーク（クラスタ）から構成され、プロセッサエレメント60をゲートウェイとして外部LAN1を介してワークステーション（WS）に接続されている。

【0092】プロセッサエレメント60にはポート・PE対応表62を備えており、内部のプロセッサエレメントで実行されるプロセスに対応する、プロセッサの物理アドレスを管理している。従って外部LANに接続された例えばワークステーションから見れば、並列システムの中のプロセッサエレメントの物理アドレスを宛先としなくてもよいことになる。

【0093】すなわち並列システム40の中のゲートウェイであるプロセッサエレメント60の物理アドレスを宛先にすれば、並列システム40の中でプロセスがどのプロセスで実行されているかを問題にしないで済むことになり、外部LANからのプロセス間通信では並列システム40の中のプロセッサのIPアドレスは一つでよいことになる。

【0094】同様に並列システム40の内部でのクラスタ間のプロセス間通信でも、プロセッサのIPアドレスは一つでよいことになる。また同様に並列システム40の内部でのクラスタ内のプロセス間通信でも、プロセッサのIPアドレスは一つでよいことになる。

【0095】図14に内部ネットワークに接続される並列システムの構成図を示す。図14では、本発明の通信宛先管理装置をホストコンピュータに適用し、プロセスを行うプロセッサエレメント（PE）を内部ネットワークを介して接続している。図14の62と64は、それぞれポート・PE対応表と内部LANデータリンクである。

【0096】尚、本発明の通信宛先管理装置は一つのホストコンピュータあるいは一つのプロセッサエレメントを兼ねたゲートウェイであってもよいし、ゲートウェイ専用装置であってもよいことは言うまでもない。図15に複数クラスタを接続するハブ（HUB）を用いた通信宛先管理システムの構成図を示す。

【0097】また外部ネットワーク1に他の並列システムあるいは分散システムを接続して大規模な並列システムあるいは分散システムを構成することができる。それぞれの並列システムもしくは分散システムでは、内部で

ネットワークを構成しており、外部ネットワークを介してそれぞれのゲートウェイで接続される。図16に大規模分散処理における通信宛先管理システム構成図を示す。

【0098】図16ではプロセスAからプロセスBへの通信の経路を示した。また図16のように発信元（プロセスA側）の並列システムでも本発明のゲートウェイをインストールすれば、プロセスBからプロセスAへの通信にも本発明の通信宛先管理が適用可能になり、大規模で通信効率の良い並列システムもしくは分散システムを構築できる。

【0099】また図16のように外部ネットワーク接続されるゲートウェイに本発明の通信宛先管理の方法および装置を適用し、内部ネットワークで接続される並列システム内では、それぞれのシステムで独自のプロトコルに基づくプロセス間通信を行うこともできる。

【0100】外部ネットワーク上ではTCP/IPなどの一般プロトコルに基づく通信を行い、それぞれの内部ネットワーク上のプロセス間通信では、内部で固有の宛先プロセスのアドレスを用いる。これにより異機種間でのプロセス間通信を外部ネットワークを通じて行うことが可能になる。

【0101】

【発明の効果】以上の説明から明らかなように本発明によれば、下位プロトコルに属するデータリンク層で通信の宛先管理を実行することができるので、上位プロトコルに必須の処理がゲートウェイでは不要になり、ゲートウェイの負荷を低減し通信の相手先のプロセッサに負荷を分散させることが可能になる、内部ネットワークに1つのIPアドレスを割り当てることができるのでIPアドレスを増やすことなしにホストコンピュータもしくはプロセッサエレメントを増やすことが可能になると共に、並列システムもしくは分散システムを1つのシステムとして扱うことが可能になる、プロセス間通信で扱う宛先アドレスを常に自動的に更新および検索できるのでユーザは今まで通りの使用法でプロセス間通信を行える、フラグメント化に対処しているのでフラグメントが発生してもフラグメントの順序性を確保できる、ゲートウェイのみならず従来からのハブにも適用することができるので、複数クラスタの通信宛先管理が可能になり、更に大規模な並列もしくは分散システムを構築できるという様々な効果がある。

【図面の簡単な説明】

【図1】 第1の発明の原理図

【図2】 第2の発明の原理図

【図3】 第3の発明の原理図

【図4】 第4の発明の原理図

【図5】 第5の発明の原理図

【図6】 本発明実施の形態のTCP/IPデータフォーマットおよびシステム構成図

- 【図7】 ポート管理クライアントサーバ
 【図8】 フラグメントの宛先管理構成図
 【図9】 フラグメント処理フロー図
 【図10】 データリンクブリッジ構成図
 【図11】 データリンクブリッジ処理フロー図
 【図12】 複数の内部LANデータリンク接続図
 【図13】 複数クラスタの通信宛先管理システム構成図

図

- 【図14】 内部ネットワークで構成される並列システム構成図
 【図15】 ハブを用いた通信宛先管理システム構成図
 【図16】 独自プロセス間通信による通信宛先管理システム構成図
 【図17】 ISOの7層参照モデル
 【図18】 TCP/IP Internet階層モデル

- 【図19】 従来のプロセス間通信の説明図

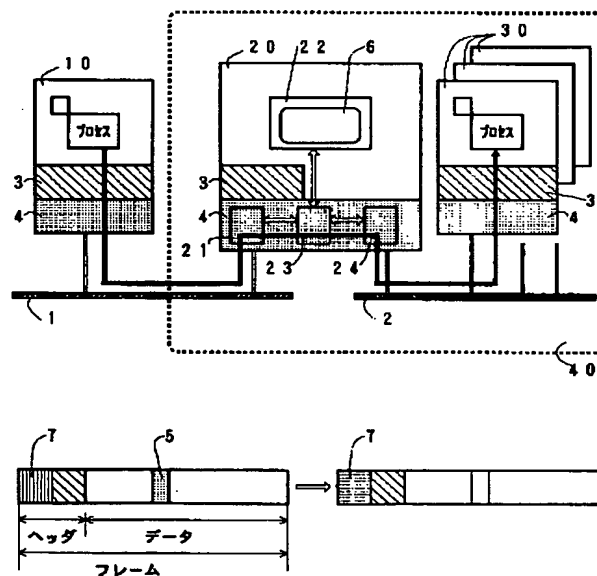
【符号の説明】

- 1 外部ネットワーク
 2 内部ネットワーク
 3 上位プロトコル
 4 下位プロトコル
 5 宛先ポート情報
 6 ポート情報に関するアドレス対応表

- 7 宛先アドレス
 10 ホストコンピュータもしくはプロセッサエレメントもしくはゲートウェイ
 11 ワークステーション
 20 ゲートウェイであり、本発明の通信宛先管理装置
 21 外部ネットワークデータリンク処理手段
 22 対応表記憶手段
 23 宛先管理手段
 24 内部ネットワークデータリンク処理手段
 25 ポート管理サーバ
 26 フラグメント記憶手段
 27 フラグメント処理手段
 28 データリンクブリッジ手段
 30 ホストコンピュータもしくはプロセッサエレメント
 40 並列あるいは分散システム
 60 ゲートウェイを兼ねたプロセッサエレメント
 61 外部LANデータリンク
 62 ポート・PE対応表
 64 内部LANデータリンク
 66 フラグメントデータバッファ
 67 IP-ID・PE対応表
 68 データリンクブリッジ
 69 ポート管理クライアント

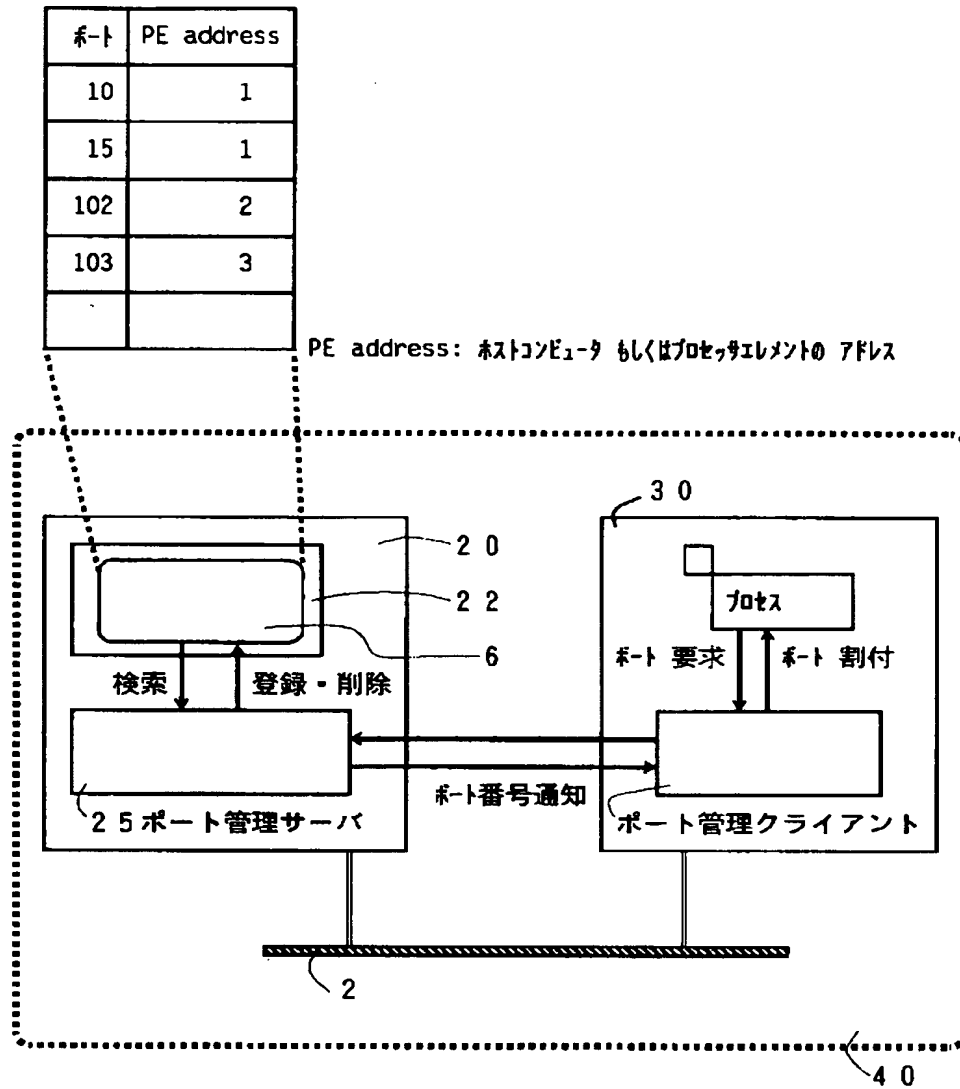
【図1】

第1の発明の原理図



【図 2】

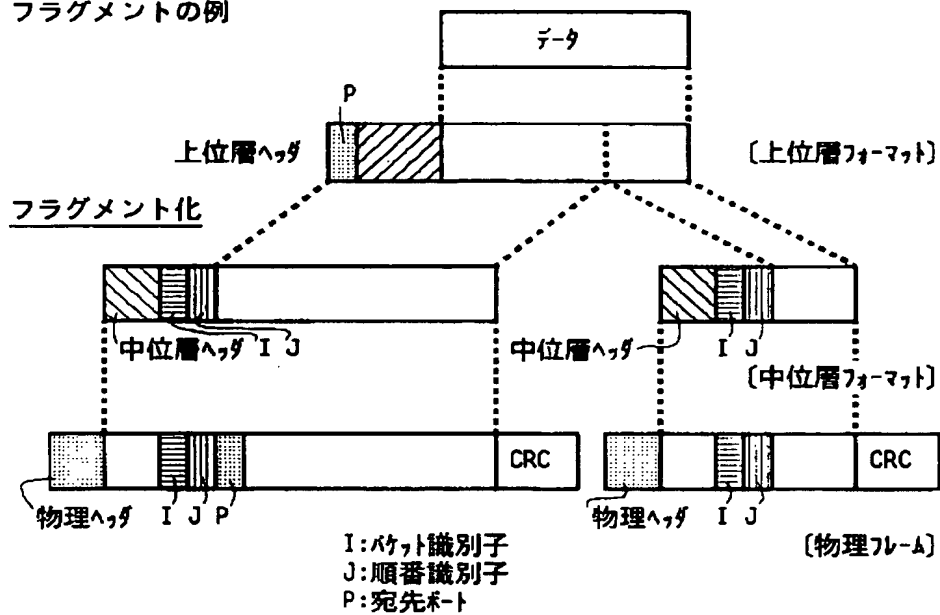
第 2 の発明の原理図



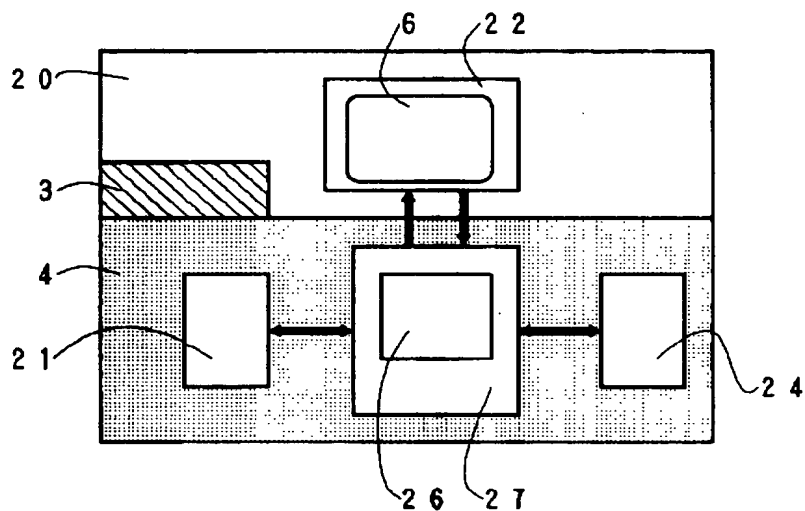
【図3】

第3の発明の原理図

(A) フラグメントの例

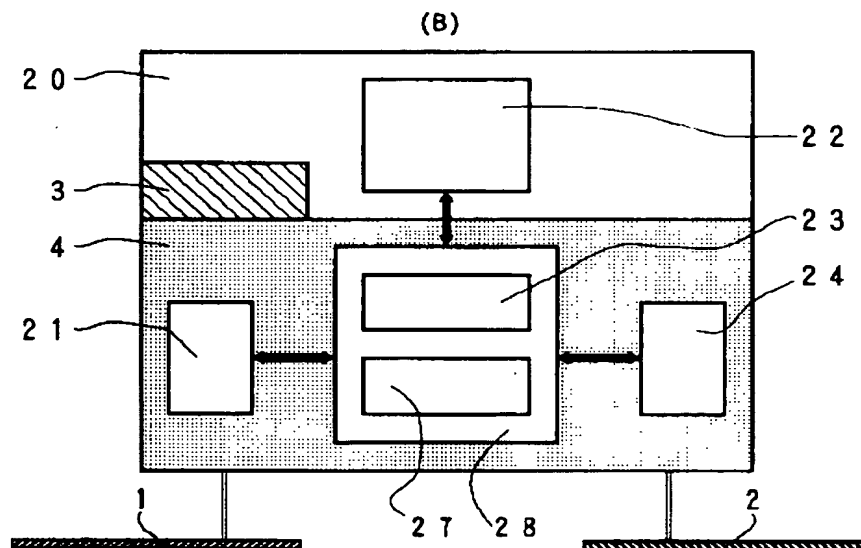
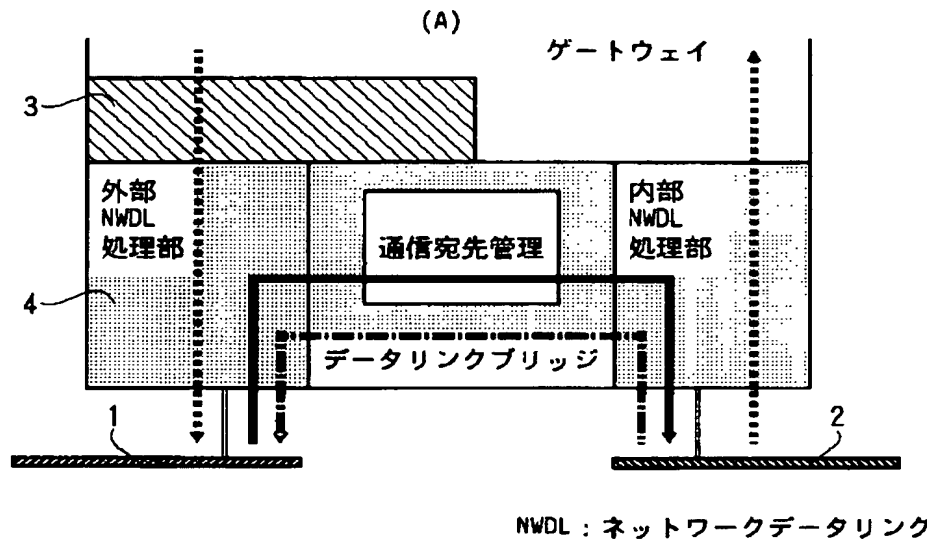


(B) フラグメントの宛先管理



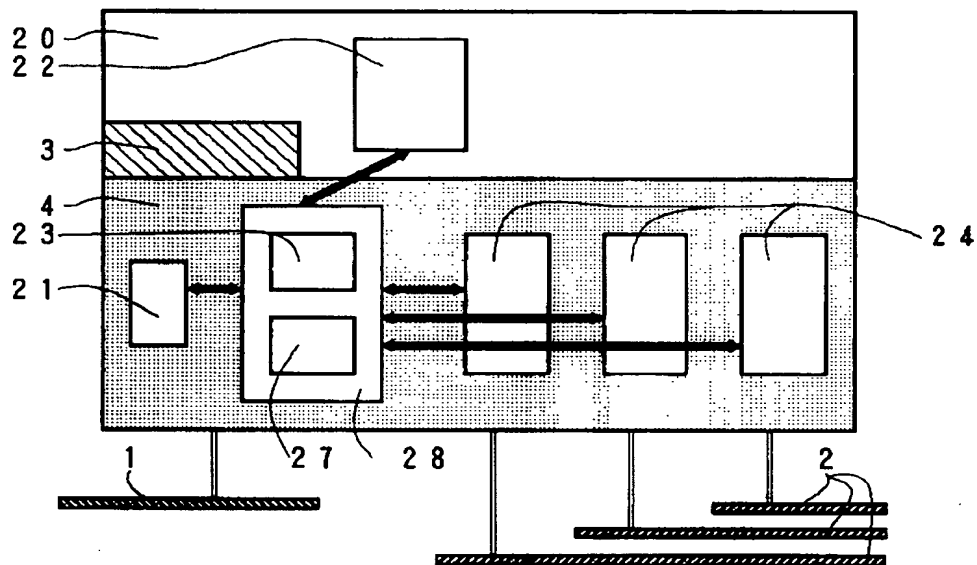
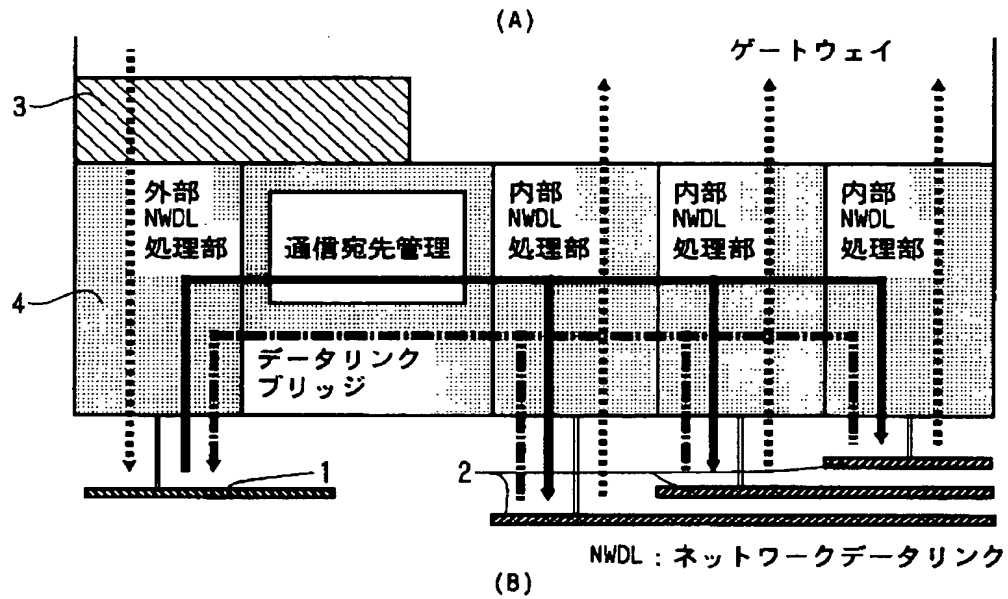
【図 4】

第 4 の発明の原理図



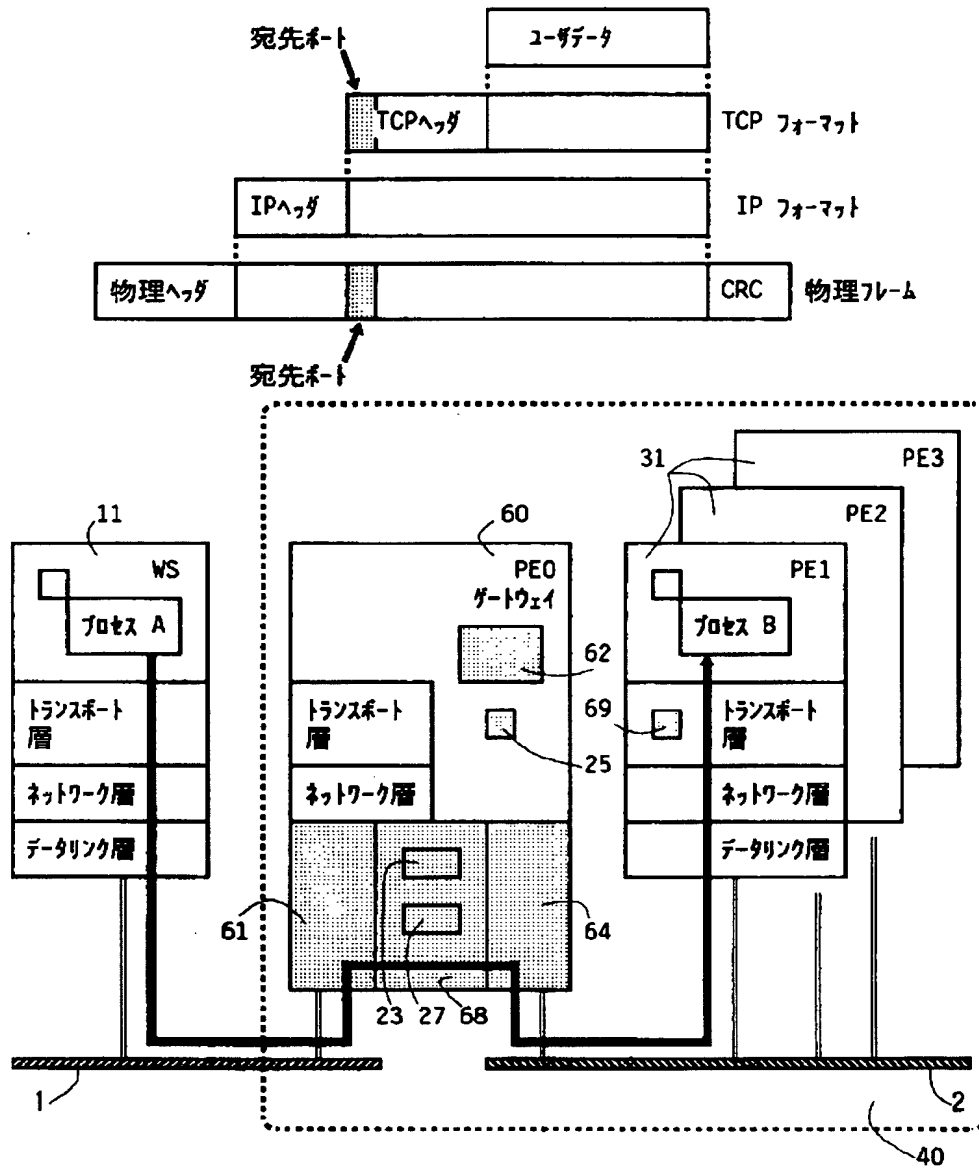
【図5】

第5の発明の原理図



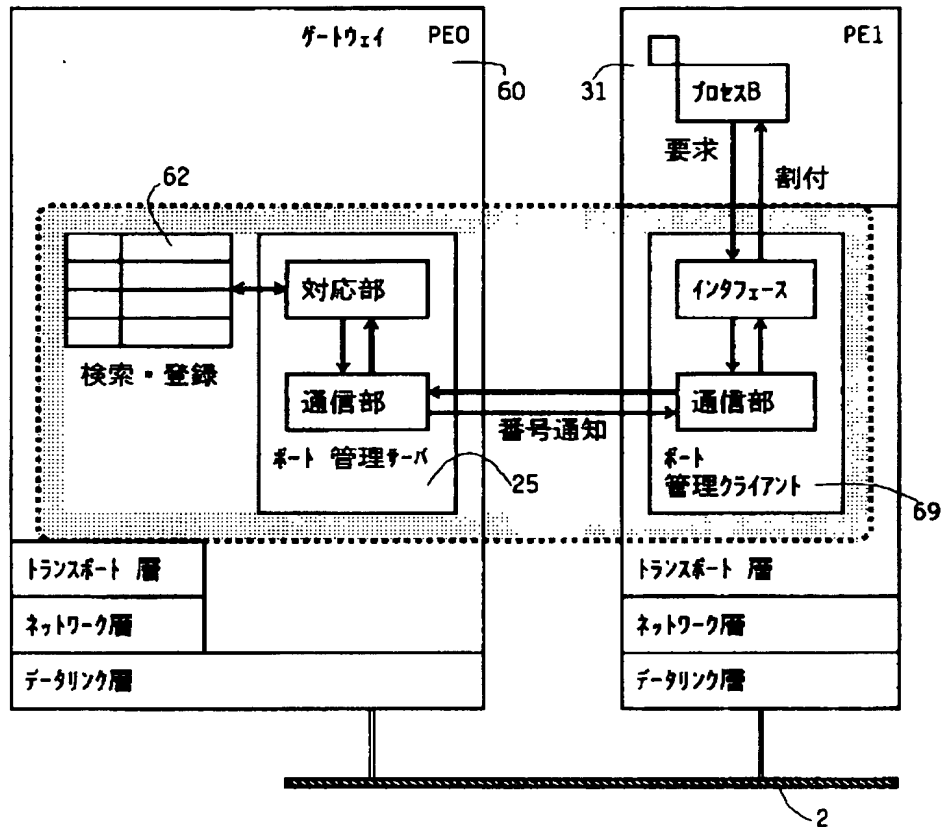
【図 6】

TCP/IPデータフォーマットおよびシステム構成図



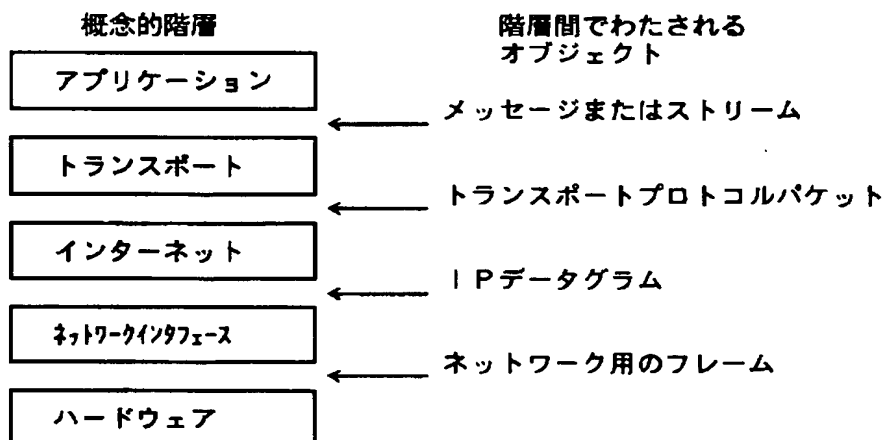
【図7】

ポート管理クライアントサーバ



【図18】

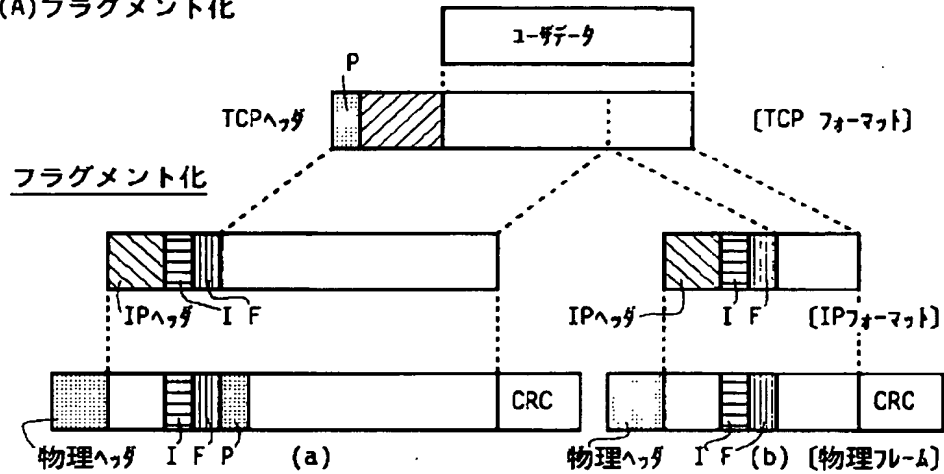
TCP/IP Internet 階層モデル



【図8】

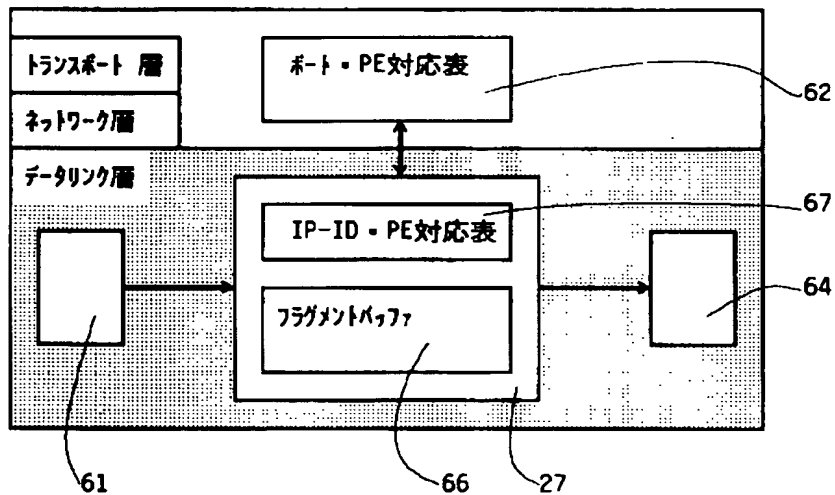
フラグメント宛先管理構成図

(A)フラグメント化



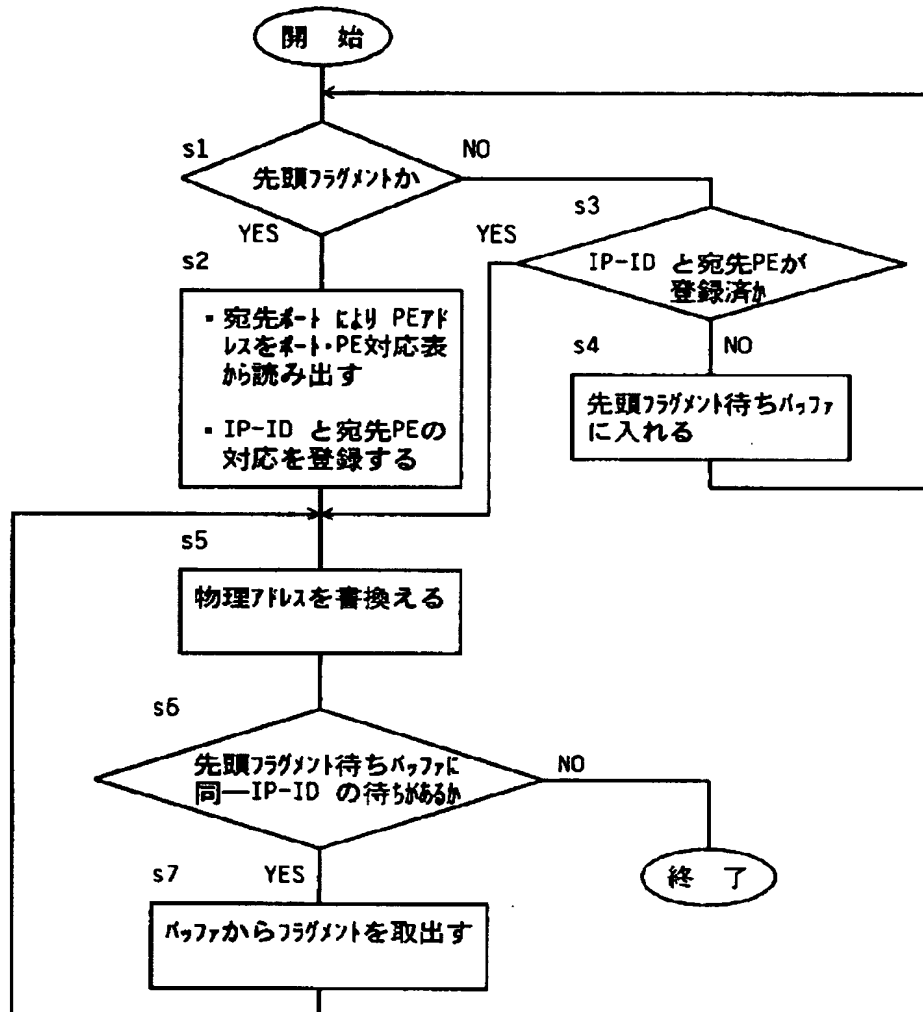
I: TCP/IP のパケット識別子である IP-IDENTIFICATION
 F: TCP/IP の順番識別子である FLAGMENT OFFSET
 P: 宛先ポート

(B)フラグメント宛先管理構成図



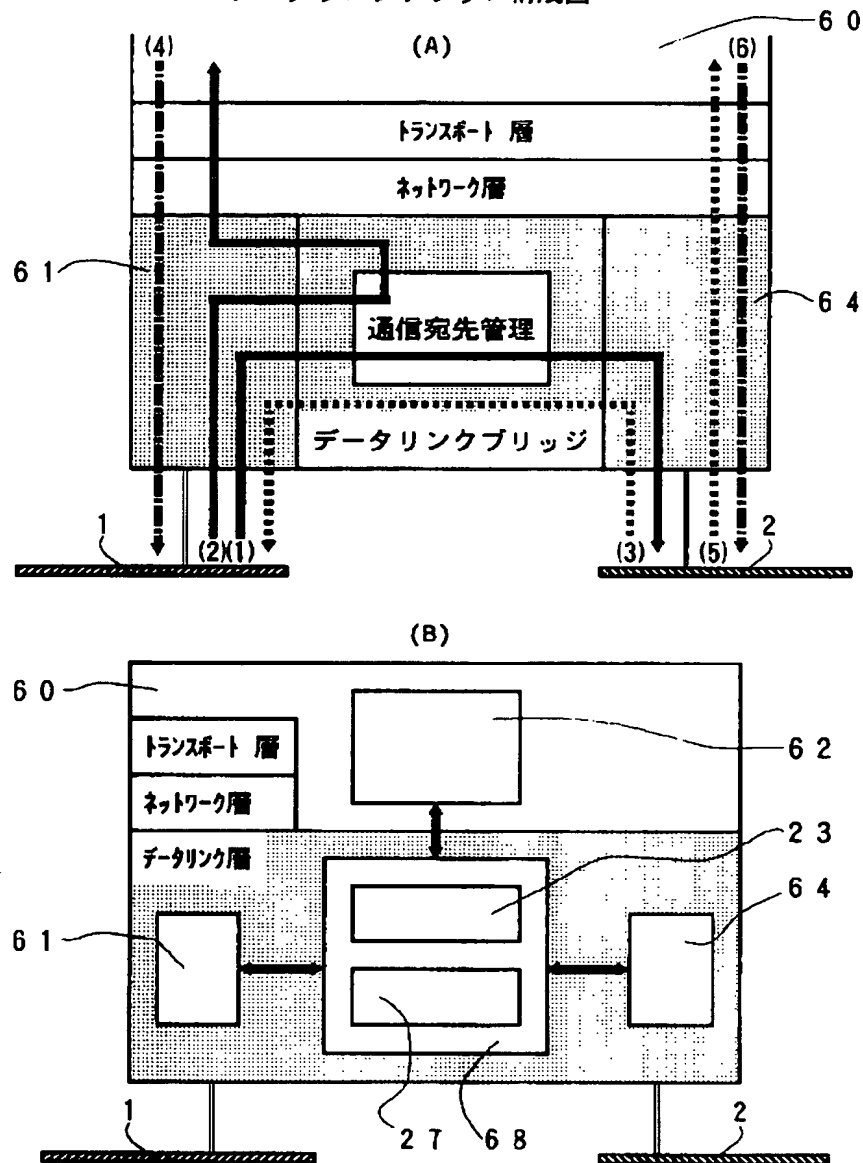
【図9】

フラグメント処理フロー図



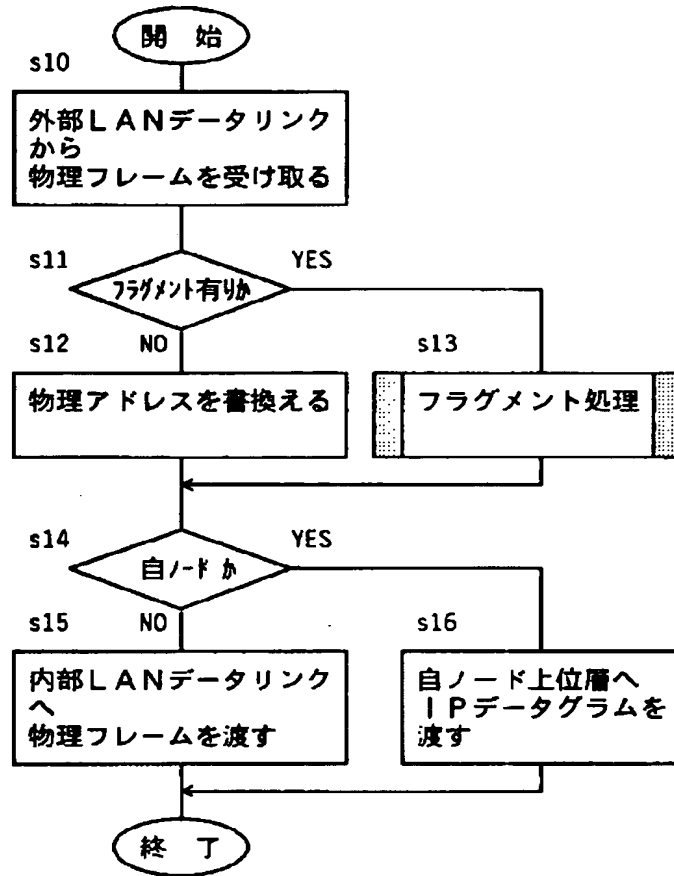
【図10】

データリンクブリッジ構成図

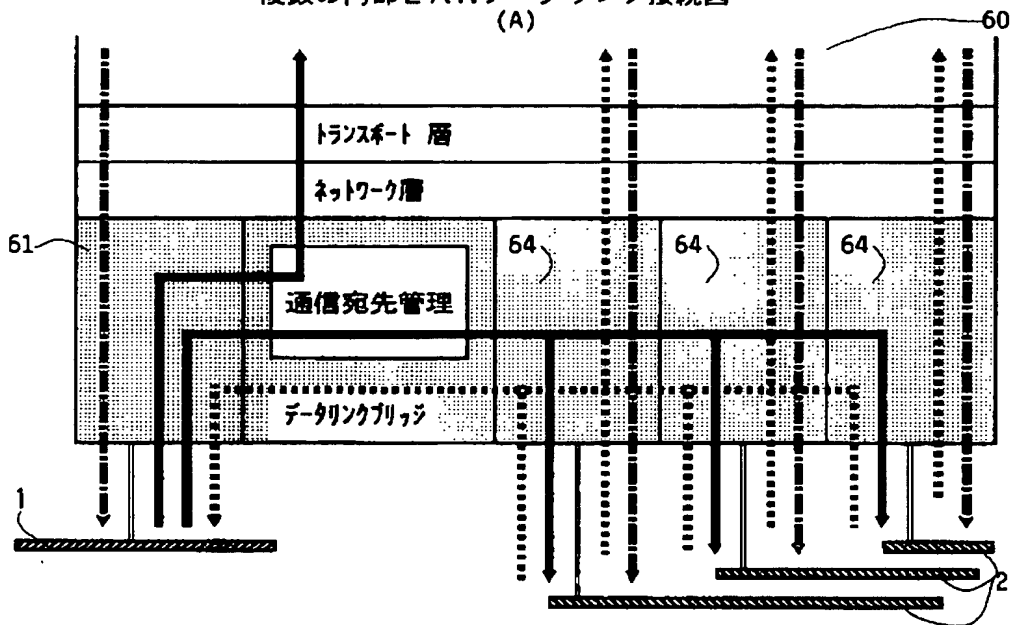


【図11】

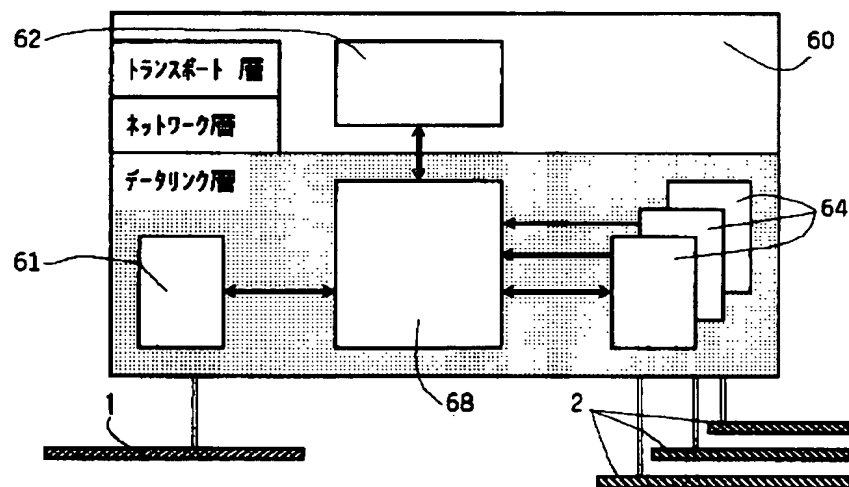
データリンクブリッジ処理フロー図



【図12】

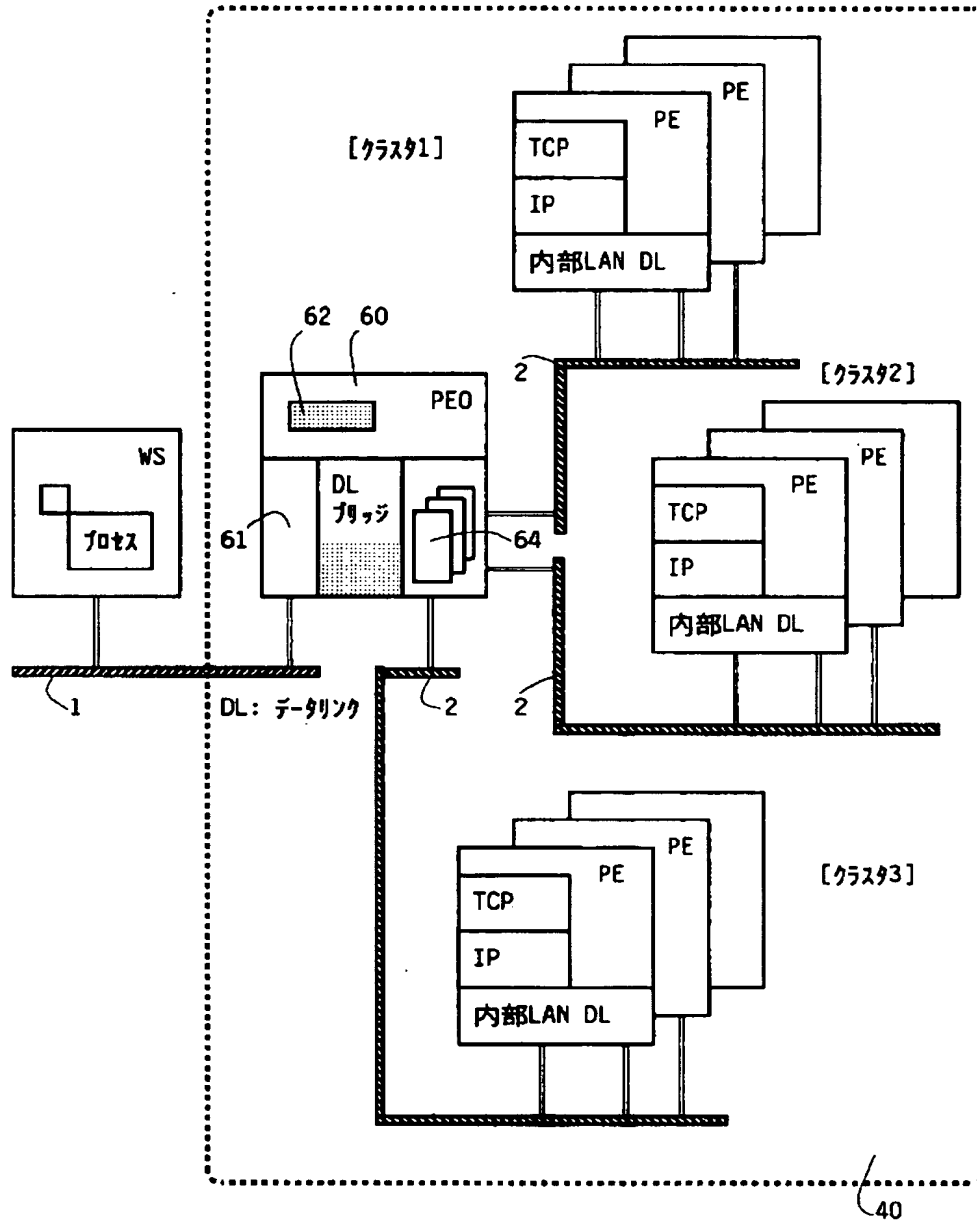
複数の内部LANデータリンク接続図
(A)

(B)



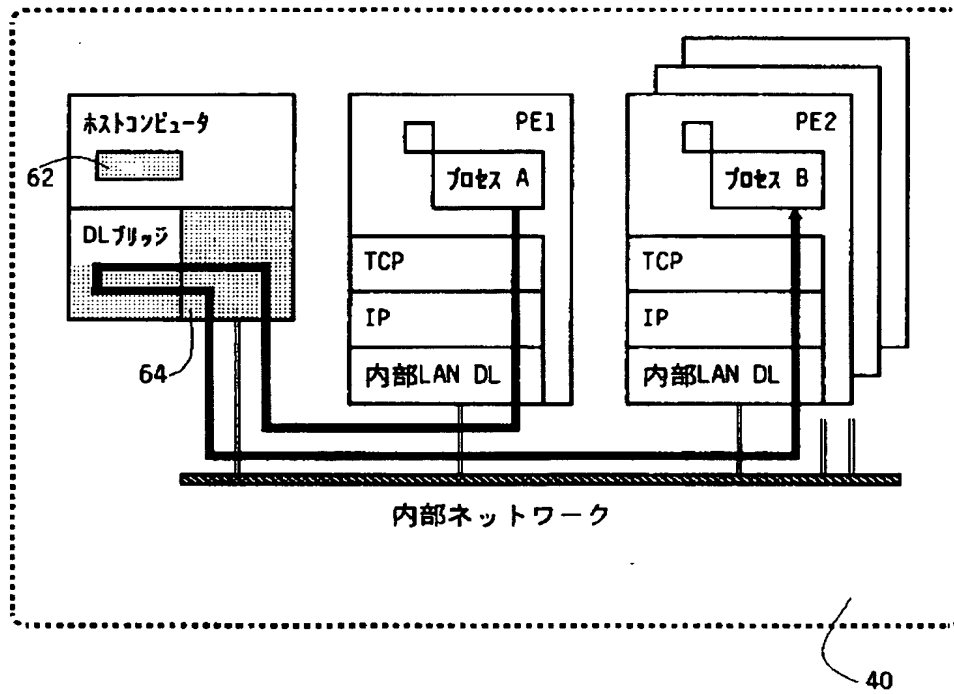
【図13】

複数クラスタの通信宛先管理システム構成図



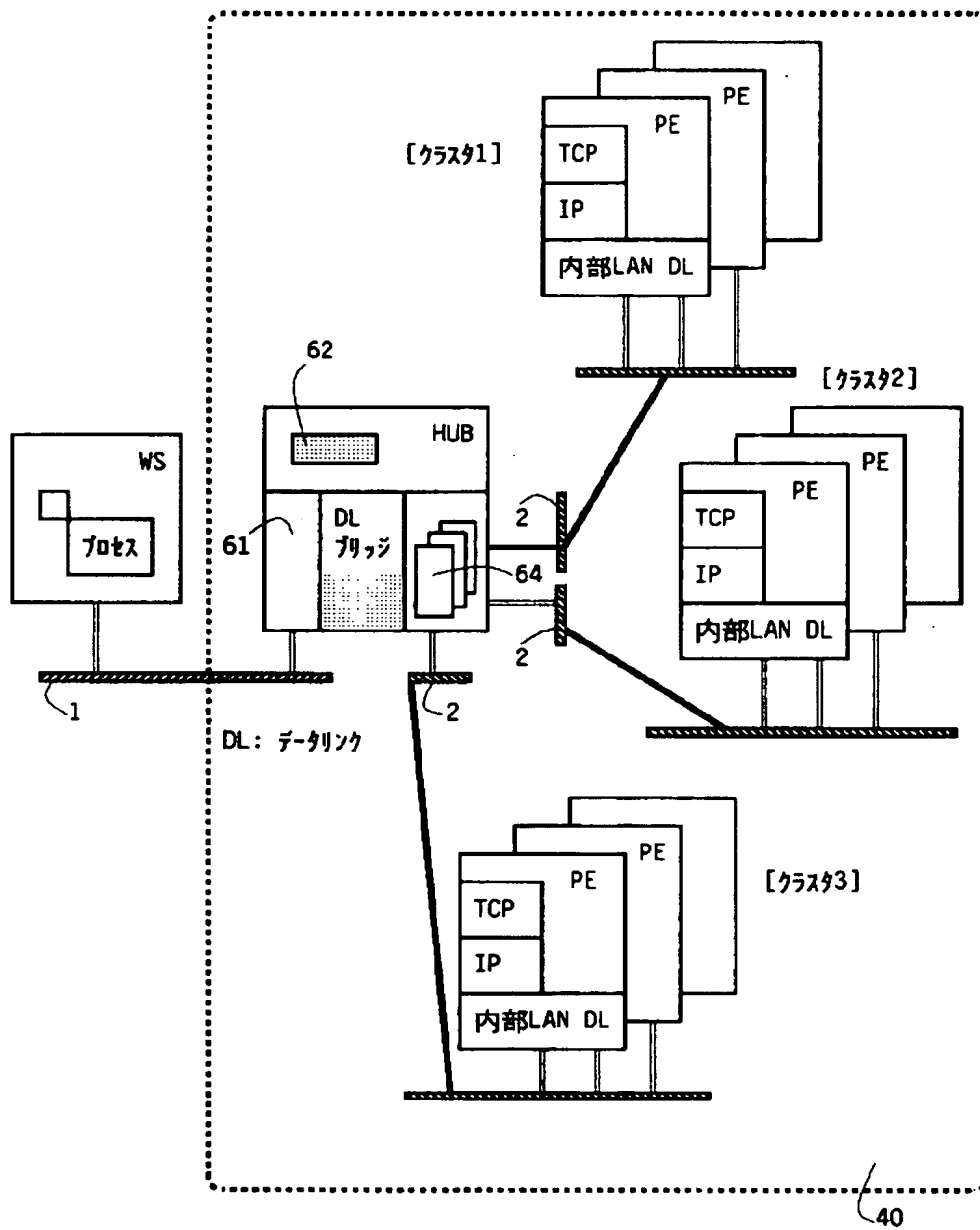
【図 1 4】

内部ネットワークで構成される並列システム構成図



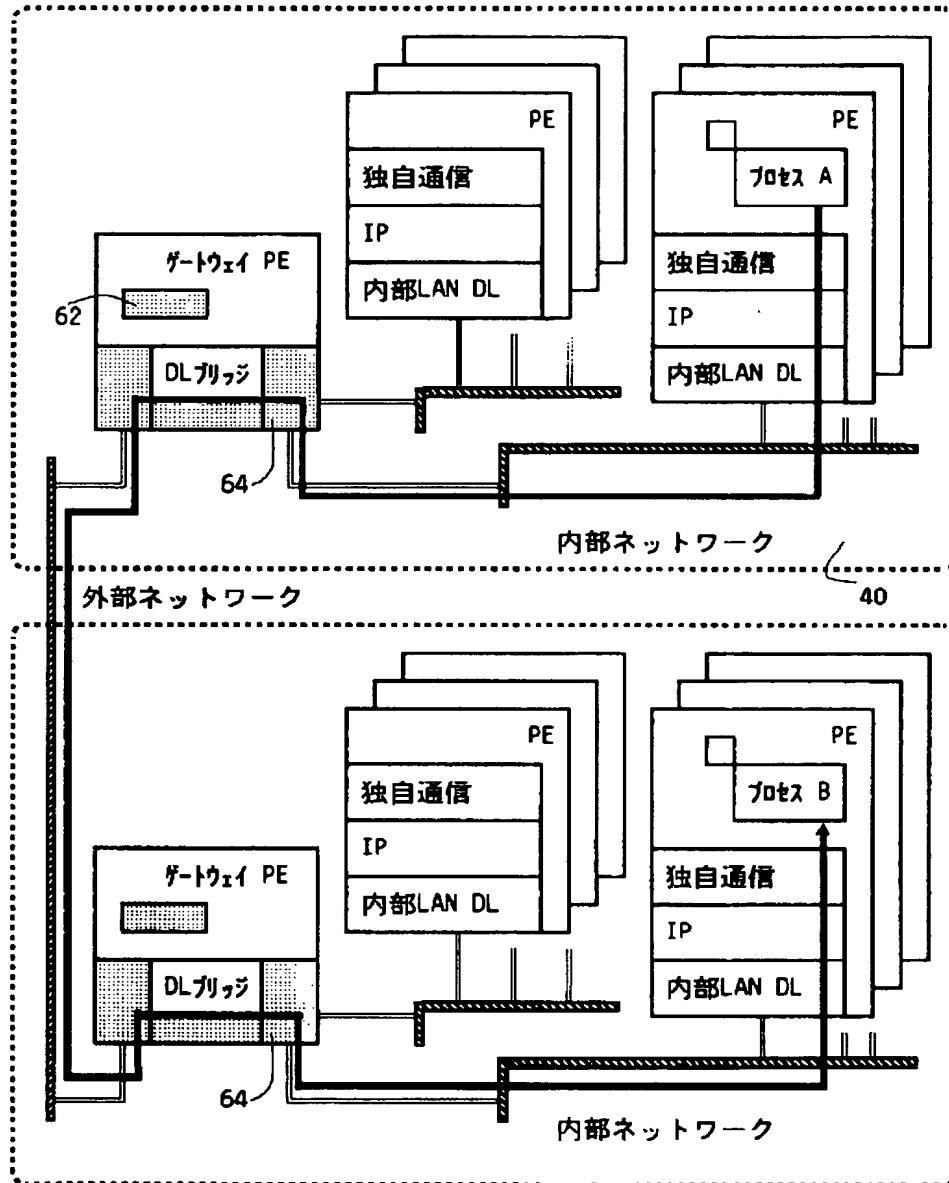
【図15】

ハブを用いた通信宛先管理システム構成図



【図16】

大規模分散処理における通信宛先管理システム構成図



【図17】

ISOの7層参照モデル

層	機能
7	アプリケーション
6	プレゼンテーション
5	セッション
4	トランスポート
3	ネットワーク
2	データリンク (ハードウェアインタフェース)
1	物理ハードウェア接続

国際標準化機構
(International Organization for Standardization (ISO))

